

Constructing low-dimensional stochastic wind models through hierarchical spatial temporal decomposition

Qiang Guo^{a,1}, D. Rajewski^b, E. Takle^b, Baskar Ganapathysubramanian^{a,*}

^a*Department of Mechanical Engineering, Iowa State University, Ames, Iowa*

^b*Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa*

Abstract

Current wind turbine simulations are usually based on simplified wind load models that assume steady wind speed and constant wind speed gradient profile. While various turbulence generating methods have been successfully used for decades, they lack of the ability to reproduce variabilities in wind dynamics, as well as inherent stochastic structures (like temporal and spatial coherences, sporadic bursts, high shear regions). It is important to incorporate these stochastic properties in analyzing and designing the next generation wind-turbines. This necessitates a more realistic parameterization of the wind that encodes location-, topography-, diurnal-, seasonal and stochastic affects. However, such a parameterization is useful in practice only if it is relatively simple (low-dimensional). Interestingly, the data to construct such models are available at various resolutions from meteorological measurements. Such meteorology data have rarely been used in wind turbine design in spite of the fact that they contain rich information about location- and topography specific wind speed data. In this work, we develop a hierarchical temporal and spatial decomposition (TSD) of large-scale meteorology data to construct a low-dimensional yet realistic stochastic wind flow model. The TSD framework is based on a two-stage model reduction: Bi-orthogonal Decomposition (BD) followed by Karhunen-Loeve expansion (KLE). We showcase this framework on data from a recent meteorological study called CWEX-11 (Crop Wind-energy EXperiment 2011). Starting from this large data set, a low dimensional stochastic parameterization is

*Corresponding author: Tel.: +1 515-294-7442; fax: +1 515-294-3261. E-mail address: baskarg@iastate.edu, URL: <http://www3.me.iastate.edu/bglab/>

¹This author is currently employed at NGC Transmission Equipment (America), Inc.

constructed using the developed framework. The resulting time-resolved stochastic model encodes the statistical behavior exhibited by the actual wind flow. Moreover, the temporal modes encode the variation of wind speed in the mean sense and resolve diurnal variations and temporal correlation while the spatial modes provide deeper insight into spatial coherence of the wind field - which is a key aspect in current wind turbine sizing, design and classification. Comparison of several important turbulent properties between the simulated wind flow and the original dataset show the utility of the framework. We envision this framework as a useful complement to existing wind simulation codes.

Keywords: stochastic wind model, reduced-order model, BD, KLE, meteorology data

1. Introduction

The reliability analysis of wind turbines requires a rigorous incorporation of the effect of randomness in the wind load. The wind load on the turbine is a stochastic process whose direction and speed depends on location, and time. In other words, “... at any instant they are distributed irregularly in space, at any point in space they fluctuate chaotically in time, and at given point and a given time they vary randomly from realization to realization.” [1] This randomness in wind conditions leads directly to the fluctuations in rotating speed of the rotor (for variable-speed wind turbine). Furthermore, randomness of wind direction causes random yaw motion of nacelle, which can further affect wind turbine’s productivity. Turbulence introduces unsteady loads on the blades. All of these effects excite structural vibrations, introduce components failures, and further reduce the lifespan of wind turbine. Thus, stochastic wind loads have a critical impact on wind turbine performance. To improve reliability, a crucial step is to have a better understanding of how they are affected by stochastic wind loads. This can be accomplished using stochastic analysis. Stochastic analysis provides strategies to evaluate the relationship between a stochastic input and various quantities of interest. In general, stochastic analysis is done by modeling the system of interest, incorporating uncertainties in system model or its inputs, and getting probability distributions of the system output. For instance, to quantify the effect of the uncertainty that is introduced by the random wind conditions, stochastic analysis (like polynomial chaos [2, 3], stochastic col-

location [4], or Monte Carlo analysis [5]) on a wind turbine model can be performed [6]. *A key ingredient to successful stochastic analysis is to have a realistic, viable models of the input variability* – in this case, stochastic wind models that encode location-, topography-, diurnal-, and seasonal affects.

Most of the past wind turbine design work was based on simplified wind load models that assume steady wind speed, constant wind speed gradient profile and constant turbulence intensity [7]. This kind of simplified wind model does not provide any insight into the effect of randomness in the wind load. They have, however, been used with some success in the context of *deterministic wind turbine* analysis [8, 9]. Subsequently, wind models that are capable of describing wind stochasticity have been developed. Gaussian and Weibull distributions have been used in approximating histograms of hourly mean wind speeds. These methods are based on parameter estimation techniques such as maximum likelihood method (MLM) (see Carta et al. [10]). While useful in representing randomness, these techniques lack the ability to describe *temporal and spatial* correlations of wind, or location-dependent characterization capabilities. This results in the development of methods that incorporate spatial variations, spatial correlation, and temporal correlation.

There have been several seminal works that deal with preserving either spatial correlation or temporal correlation. One of such examples is Veers/Sandia method [11]. The method first uses empirical coherence function and wind speed power spectral densities (PSDs) to describe the wind field. Choleski decomposition is then used to decompose the cross spectrum matrix. Velocity time-series at different locations are finally calculated with a inverse Fourier transform process. Another example is Mann’s method [12], where the wind field is described in wave vector space and then transformed into spatial domain. Recently, proper orthogonal decomposition (POD) was used [13] to simulate wind turbine inflow. In this method, stochastic wind flow is viewed as wind field snapshots at different instances. A spatial covariance matrix is decomposed to get multiple characteristic modes that are further used in constructing a stochastic wind field. By performing POD analysis at each time step, the wind flow is reproduced. Although these three methods are easy to implement and efficient enough to get random wind flows, they do not account for *both temporal correlation and spatial correlation of the wind flow at the same time*. In the software package TurbSim [14], a turbulence simulation code developed by National Renewable Energy Laboratory (NREL), Veers method is used. However, due to its limitation in describing temporal coherence, additional information about

coherent structure has to be included in the model so that the turbulent structure in atmosphere is more accurately represented [15].

The motivation for this work is the fact that no single method exists that can provide a data-driven model that seamlessly accounts for spatial variations, diurnal variations, and temporal correlations ². To this end, an efficient method that leverages the huge amount of meteorological data that is readily available will be very useful for various aspects of wind turbine analysis. This paper describes such a technique. The contributions of this work are as follows: 1) we formulate a mathematical framework that is able to accurately represent wind flow using a low-complexity yet realistic stochastic model; 2) we implement a computational framework based on the mathematical framework that extracts statistical information from meteorology data and constructs an easy-to-use model; 3) we test the model by generating synthetic stochastic wind flow and comparing several wind turbulence statistics between synthetic and original flows.

In the following sections, section 2 focuses on constructing a low-complexity model of the random wind flow. Bi-orthogonal Decomposition (BD) [17, 18] and Karhunen-Loève expansion (KLE) [5] are used in constructing the low-dimensional wind model. A numerical example uses CWEX-11 (Crop Wind-energy EXperiment 2011) data is given in section 3. The algorithmic details of the computational framework is give in section 4. The result of the analysis is shown in section 5. In order to quantitatively check the accuracy of the decomposition, several statistical properties of the synthetic stochastic wind flow and the original wind flow are compared. Finally, section 6 concludes the paper with the main findings and their implications.

²The methods discussed here are invariably parametric – where the model (and parameters) are set based on priori assumptions. For example, the probability density function of the velocity fluctuations in homogeneous turbulence is assumed to be Gaussian or sub-Gaussian [16]. Although this assumption is often valid for many physical phenomena (including wind on flat terrain), it is not necessarily true in the case of complicated terrain. The alternative approach – nonparametric methods – estimate distributions entirely based on data. Although robust and easy to implement, the accuracy of non-parametric methods strongly depend on the availability of data samples. In this paper, a nonparametric method, kernel density estimation (KDE), is used to describe the stochasticity of the turbulence. It is worth noting that the applicability of the framework developed in this context should not depend on the choice of density estimation techniques.

2. Mathematical Framework

A stochastic wind can be denoted by $\mathbf{v}(\mathbf{x}, t, \boldsymbol{\xi})$, where $\mathbf{x} = (x_1, x_2, x_3) = (x, y, z)$ is the three-dimensional coordinates with x , y , and z represents along-wind, transverse, and vertical axis respectively, t denotes time, and $\boldsymbol{\xi}$ represents stochastic variability. Wind velocity $\mathbf{v} = (v_1, v_2, v_3) = (u, v, w)$ consists of three wind speed components along the three spatial axes.

The analysis is clearer when the mean component $\bar{\mathbf{v}}(\mathbf{x})$ of the velocity data is removed so that only the fluctuation components $\mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi})$ are left, i.e.

$$\mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}) = \mathbf{v}(\mathbf{x}, t, \boldsymbol{\xi}) - \bar{\mathbf{v}}(\mathbf{x}), . \quad (1)$$

The mean component is defined as

$$\bar{\mathbf{v}}(\mathbf{x}) = \frac{1}{|T|} \int_T \langle \mathbf{v}(\mathbf{x}, t, \boldsymbol{\xi}) \rangle dt. \quad (2)$$

where $\langle \cdot \rangle$ denotes the average in the stochastic domain, $|T|$ is the span of the temporal domain, $\bar{\mathbf{v}}$ is the ensemble average.

The goal of this work is to construct a simple stochastic model that encodes temporal and spatial covariance, and preserve all the statistical information, such as spatial coherence and wind speed power spectral density (PSD), present in the collected meteorological data. We look for a model that has the form

$$\mathbf{u} = \sum K_{ijk} \mathbf{X}_i(\mathbf{x}) T_j(t) \boldsymbol{\xi}_k \quad (3)$$

where K_{ijk} are coefficients, the deterministic functions \mathbf{X}_i track spatial correlations, T_j track temporal correlations and the random variables $\boldsymbol{\xi}_k$ encode the inherent variability. *The goal becomes to find the simplest possible representation that still encodes all the required information.*

The main idea of the next section is to formulate a mathematical strategy of representing this meteorology data in terms of the smallest possible number of terms in Eqn. 3, by optimally designing \mathbf{X} , T , and $\boldsymbol{\xi}$. Noting that the data contains spatial, temporal and stochastic variabilities, we solve this problem in two stages. In the first stage, we decompose the data into temporal ($T(t)$) and coupled spatial-stochastic ($\boldsymbol{\Phi}(\mathbf{x}, \boldsymbol{\xi})$) parts through the concept of Bi-orthogonal Decomposition; in the second part we decompose the spatial-stochastic part into spatial ($\mathbf{X}(\mathbf{x})$) and stochastic ($\boldsymbol{\xi}$) components using the concept of the Karhunen-Loève decomposition.

2.1. Stage 1: A low-dimensional representation via Bio-orthogonal decomposition

In the first stage, we are looking for a minimal representation of the data in the form

$$\mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}) \approx \sum_{i=1}^M K_i \boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi}) T_i(t). \quad (4)$$

where $\boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi})$ are stochastic spatial modes and $T_i(t)$ are temporal modes.³ Our goal becomes searching for the best choices for $\boldsymbol{\Phi}_i$ and T_i such that the decomposition uses the least number of terms M that will give us an accurate representation of $\mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi})$. We pose this as an optimization problem. To do so, we define the error in this representation and design $\boldsymbol{\Phi}_i$ and T_i that minimize this error. The error, denoted by ε , is defined as the low-complexity stochastic model subtracted from the true data

$$\varepsilon(\mathbf{x}, t, \boldsymbol{\xi}) = \mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}) - \sum_{i=1}^M K_i \boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi}) T_i(t). \quad (5)$$

Approximation theory suggest that the best choice for the functions $\boldsymbol{\Phi}_i$ and T_i (we will also call them modes) is when they are orthogonal to each other [20]. Thus, we set T_i , $i = 1, \dots, M$ to be orthogonal to each other in the time domain and $\boldsymbol{\Phi}_i$, $i = 1, \dots, M$ to be weakly orthogonal in spatial domain. Mathematically, this is denoted in terms of the inner products:

$$\langle T_i, T_j \rangle_T = \int_T T_i(t) T_j(t) dt = \delta_{ij} \quad (6)$$

and

$$\langle \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j \rangle_X = \int_X \overline{\boldsymbol{\Phi}_i} \cdot \overline{\boldsymbol{\Phi}_j} d\mathbf{x} = \delta_{ij}, \quad (7)$$

³A general representation will be of the form $\mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}) = \sum_{i,j} K_{ij} \boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi}) T_j(t)$, where i and j are independent indices. Consequently, the expression on the right hand side has an dramatically large number of terms. This is handled utilizing the Schmidt decomposition theorem [19], which states that – any representation of a tensor product space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ can be expressed as linear combination of tensor product of basis functions $\boldsymbol{\Phi}_i \otimes \Psi_i$, where $\boldsymbol{\Phi}_i \in \mathcal{H}_1$, $\Psi_i \in \mathcal{H}_2$. As a result, the representation can be reduced to Eqn. 4.

where $\overline{\Phi_i}$ denotes the expectation of the spatial-stochastic mode, i.e.

$$\overline{\Phi_i(\mathbf{x})} = \int \Phi_i(\mathbf{x}, \boldsymbol{\xi}) W(\boldsymbol{\xi}) d\boldsymbol{\xi}, \quad (8)$$

and $W(\boldsymbol{\xi})$ is the multivariate joint probability density of random variables in the set $\boldsymbol{\xi}$.

Note that the error $\boldsymbol{\varepsilon}$ itself is an random field. We construct an associated scalar value with this random field to accomplish subsequent optimization. To this end, an error functional \mathcal{E} is defined as the norm of $\boldsymbol{\varepsilon}$, i.e.

$$\mathcal{E} = \int_T \langle \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \rangle_X dt. \quad (9)$$

The error-functional is simply the inner product (i.e. an average) of the field over space, time and stochastic dimensions. Note that the error functional depends on the choice of functions $T_i(t)$ and $\Phi_i(\mathbf{x}, \boldsymbol{\xi})$, i.e. $\mathcal{E}[T_1, \dots, T_M, \Phi_1, \dots, \Phi_M]$.

We now search for temporal functions that minimizes this error functional. This is accomplished by applying the calculus of variations and solving the associated Euler-Lagrange equations. We provide full details of the derivation in Appendix A. This optimization reduces to the solution of an eigenvalue problem, where the eigenfunctions give the temporal modes:

$$\mu_i T_i(t) = \int_T C(t, t') T_i(t') dt', \quad (10)$$

where $C(t, t')$ is the temporal covariance matrix of the meteorology data, and μ_i are the eigenvalues of this temporal covariance matrix. $C(t, t')$ is constructed by taking the inner product of the data in the spatial domain, i.e

$$C(t, t') = \langle \mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}), \mathbf{u}(\mathbf{x}, t', \boldsymbol{\xi}) \rangle_X. \quad (11)$$

Once μ_i and T_i are solved for, the spatial-stochastic functions are calculated using

$$\Phi_i(\mathbf{x}, \boldsymbol{\xi}) = \frac{1}{\sqrt{\mu_i}} \langle \mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}), T_i(t) \rangle_T. \quad (12)$$

The first M (usually $M \sim 3 - 6$) eigenvalues and eigenfunctions usually represent the data exceedingly well [4]. Thus, the first stage of the decomposition of the wind data results in representation involving temporal functions

T_i and spatial-stochastic functions $\Phi(\mathbf{x}, \xi)$. Eqn. 4 now becomes

$$\mathbf{u}(\mathbf{x}, t, \xi) = \sum_{i=1}^M \mathbf{a}_i(\mathbf{x}, \xi) T_i(t), \quad (13)$$

where

$$\mathbf{a}_i(\mathbf{x}, \xi) = K_i \Phi_i(\mathbf{x}, \xi) = \sqrt{\mu_i} \Phi_i(\mathbf{x}, \xi). \quad (14)$$

are spatial-stochastic modes. The second stage of the framework involves decomposing the spatial-stochastic functions $\mathbf{a}_i(\mathbf{x}, \xi)$ into spatial functions $\mathbf{X}(\mathbf{x})$ and independent random variables ξ .

Remark 1. *We chose to decompose the data into spatial-stochastic and temporal parts in the first stage of the decomposition. This decomposition is one of three possible decomposition choices. These choices are enumerated in Table 1 where \mathbb{X} , \mathbb{T} , and Σ denote the spatial, temporal, and stochastic domains respectively. Such decompositions have been explored in other works. For instance, Venturi et al. [17, 18] investigated Type 1 decomposition. The decomposition suggested by Type 2 can be achieved by using generalized polynomial chaos [2, 3]. Mathelin et al. [21] modeled uncertain cylinder wake using a Type 3 decomposition. It can be shown that Type 1 and Type 3 decompositions result in identical results. \square*

Table 1: Choices for Bi-orthogonal Decomposition.

Type	Space 1	Space 2
1	$\mathbb{X} \times \Sigma$	\mathbb{T}
2	$\mathbb{X} \times \mathbb{T}$	Σ
3	$\mathbb{T} \times \Sigma$	\mathbb{X}

Remark 2. *The choice of the inner products (temporal Eqn. 6 and spatial-stochastic Eqn. 7) affect the properties of the decomposition. In particular, there are several different ways in which we can define an inner product over the spatial-stochastic modes (i.e. different ways to average over space and stochastic dimensions). These include the following possibilities:*

$$\langle \Phi_i, \Phi_j \rangle_0 = \int_{\mathbf{X}} \overline{\Phi_i} \cdot \overline{\Phi_j} d\mathbf{x},$$

$$\begin{aligned}\langle \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j \rangle_1 &= \int_{\mathbf{X}} \overline{\boldsymbol{\Phi}_i \cdot \boldsymbol{\Phi}_j} d\mathbf{x}, \\ \langle \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j \rangle_2 &= \int_{\mathbf{X}} \overline{\boldsymbol{\Phi}_i \cdot \boldsymbol{\Phi}_j} - \overline{\boldsymbol{\Phi}_i} \cdot \overline{\boldsymbol{\Phi}_j} d\mathbf{x},\end{aligned}$$

The first inner-product (denoted by $\langle \cdot \rangle_0$) is a spatial integral of the product of expected values, while the second inner-product (denoted by $\langle \cdot \rangle_1$) is the expectation of the spatial integral. It can be shown that by taking inner products, $\langle \cdot \rangle_h$, of type $h = 0, 1, 2$, we obtain optimal representations with respect to mean, second-order moment, and standard deviation of the data, respectively. We have chosen to focus on representation that is optimal in the mean sense [17]. \square

2.2. Stage 2: Karhunen-Loève expansion of spatial stochastic modes

In this stage, we decompose the spatial-stochastic functions $\boldsymbol{a}_i(\mathbf{x}, \boldsymbol{\xi})$ into a spatial part and a set of uncorrelated random variables. First, the mean of spatial-stochastic functions $\bar{\boldsymbol{a}}(\mathbf{x})$ are removed so that only the fluctuation component is left for analysis.

$$\boldsymbol{\alpha}(\mathbf{x}, \boldsymbol{\xi}) = \boldsymbol{a}(\mathbf{x}, \boldsymbol{\xi}) - \bar{\boldsymbol{a}}(\mathbf{x}). \quad (15)$$

Following the rational of Bi-orthogonal Decomposition (Eqn. 4), our goal is to decompose $\boldsymbol{\alpha}(\mathbf{x}, \boldsymbol{\xi})$ into a minimal set of linear combination of deterministic spatial functions and uncorrelated random variables, that is

$$\boldsymbol{\alpha}(\mathbf{x}, \boldsymbol{\xi}) \approx \sum_{i=1}^N C_i \xi_i \mathbf{X}_i(\mathbf{x}). \quad (16)$$

where C_i are coefficients of the expansion, ξ_i is a set of uncorrelated random variables, \mathbf{X}_i are deterministic spatial functions. We pose this decomposition problem as an optimization problem, where the optimization problem is to minimize the error. The representation error is defined as

$$\boldsymbol{\varepsilon} = \boldsymbol{\alpha}(\mathbf{x}, \boldsymbol{\xi}) - \sum_{i=1}^N C_i \xi_i \mathbf{X}_i(\mathbf{x}) \quad (17)$$

This is a standard formulation of the Karhunen-Loève expansion. The goal of KLE is to find the optimal choice for functions \mathbf{X}_i such that the representation error is minimized with a finite number (N) of expansion terms. We

briefly describe the mathematical framework of KLE below [5]: The representation error is converted into a cost-functional for optimization by simply considering the mean-square error (i.e. the inner-product)

$$\mathcal{E}^2 = \int_{\mathbf{X}} \bar{\epsilon}^2 d\mathbf{x}. \quad (18)$$

Minimization of the error-functional results in an eigenvalue problem, whose eigenfunctions are the desired spatial functions \mathbf{X}_i

$$\int_{\mathbf{X}} R(\mathbf{x}_1, \mathbf{x}_2) \mathbf{X}_i(\mathbf{x}_2) d\mathbf{x}_1 = C_i^2 \mathbf{X}_i(\mathbf{x}_2), \quad (19)$$

where $R(\mathbf{x}_1, \mathbf{x}_2)$ is the covariance kernel constructed from the spatial-stochastic functions $\alpha(\mathbf{x}, \xi)$. We denote $C_i^2 = \lambda_i$ so that λ_i and $\mathbf{X}_i(\mathbf{x})$ are the eigenvalues and the eigenvectors of the covariance kernel. Eqn. 16 becomes

$$\alpha(\mathbf{x}, \xi) \approx \sum_{i=1}^N \sqrt{\lambda_i} \xi_i \mathbf{X}_i(\mathbf{x}). \quad (20)$$

The final stage of the formulation is to identify the probability distributions of the uncorrelated random variables ξ_i . Note that we have the same number of realizations of ξ_i as the number of random samples of the stochastic wind. Each realization is computed by inverting Eqn. 20 for ξ_i :

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \int_{\mathbf{X}} \alpha \mathbf{X}_i(\mathbf{x}) d\mathbf{x}. \quad (21)$$

The choice of technique to construct the probability distribution of ξ_i given a finite number of observations of ξ_i is crucial. We utilize Kernel Density Estimation (KDE) methods [22] to construct the probability distributions of ξ_i in a non-parametric way.

Remark 3. *In order to solve the eigen-problem (Eqn. 19), the covariance kernel must first be calculated from the data. The Wiener-Khinchin theorem allows computing the covariance in a very efficient way. Representing the spatial-stochastic data as a matrix I , the covariance, R , is computed as:*

$$R = \mathcal{F}^{-1}(\mathcal{F}(I) \times \mathcal{F}(I)'). \quad (22)$$

where $\mathcal{F}(I)'$ is the complex conjugate of Fourier transform of I , and the diagonal entries of R contains the covariance. Numerical details for solving this generalized eigenvalue problem is included in Appendix B and a software available [23]. \square

Remark 4. In KDE, the PDF of a random variable ξ is estimated as

$$p(\xi) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\xi - \xi_i}{h}\right), \quad (23)$$

where $K(\cdot)$ is the kernel which is a symmetric function that integrates to one, and $h > 0$ is a smoothing parameter called the bandwidth. Common choices for $K(\cdot)$ are the multivariate Gaussian density function. The choice of the bandwidth is critical since a small values of h result in the estimated density with many 'wiggles', while large values of h result in very smooth estimations that do not represent the local distributions. Several approaches have been developed to chose proper h values. A good review on bandwidth selection can be found in [24]. We advocate using a simple formula based on Silverman's rule [25]. The optimal choice for h according to Silverman's rule is given by

$$h = \left(\frac{4\hat{\sigma}}{3N}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma}n^{-\frac{1}{5}}, \quad (24)$$

where $\hat{\sigma}$ is the standard deviation of the samples. \square

Based on the above three techniques, we are now able to construct a space-time decomposition of the wind field snapshots and preserve the spatial and temporal correlations. In next section, we look at a numerical example that illustrates the power of this framework.

3. Numerical Example: CWEX-11

We focus on constructing realistic low-complexity models from experimental data. We illustrate the methodology based on a recent meteorological experiment named CEWX-11. In this context, the CWEX-11 data is ideal to test the framework and demonstrate the advantages of the framework. It is straightforward to extend this analysis to any other meteorological dataset

that expands the scale of measurements of the CWEX-11 experiment.⁴

CWEX-11 is a collaborative experiment (Iowa State University (ISU) and the University of Colorado (UC), assisted by the National Center for Atmospheric Research (NCAR)). CWEX-11 and its 2010 counterpart (CWEX-10) in the Crop Wind-energy EXperiment, address observational evidence for the interaction between large wind farms, crop agriculture, and surface-layer, boundary-layer, and mesoscale meteorology (Rajewski et al. 2012) [26]. In the experiment, a surface flux station was installed to the south of a wind turbine, which makes the station measure upstream inflow of the wind turbine due to the fact that predominant summer winds in Iowa originate from south to slightly south-east. The surface station was equipped with a CSAT3 sonic anemometer that was located at height 4.5m and an RMY propeller and vane anemometer that was located at height 10.0m. The former measured wind speed in 3 directions at 20 Hz whereas the latter gave wind speed amplitude and its direction at 1 Hz. The experiment started on June 29 and lasted for 48 days.⁵

A schematic of the experiment is shown in Fig. 1. Note that the figure is for illustrative purpose and is not drawn to scale. In this experiment, only wind speed magnitude was analyzed, but it is straightforward to perform analysis on any one of the three components of turbulence.

Ideally, we would like to have multiple wind field snapshots taken at regular intervals on a vertical plane that is perpendicular to the rotor. However, due to the constraints of the experiment, only two time-series were measured at heights 4.5m and 10.0m. Linear interpolation of the two time-series at 18 spatial points on z direction were performed to simulate more vertical measurements. The wind snapshots are constructed by using *Taylor's frozen turbulence hypothesis* [27]. Specifically, all measurements in the resulting 20 time-series over certain interval were treated as taken at the same instant in the interval. This provides us along-wind (x direction) measurements. Fi-

⁴The CWEX-11 data measures the wind profile at two locations: 4.5 meters and 10 meters. Obviously, this data does not describe the wind speed profile at hub height. Additionally, measurements at only two different heights are certainly not enough to capture all the spatial stochasticity in the wind.

⁵Excluding the days that mostly have opposite wind direction and the days with sudden rain, only data for 28 days that had the desired weather condition and wind direction were used in the analysis.

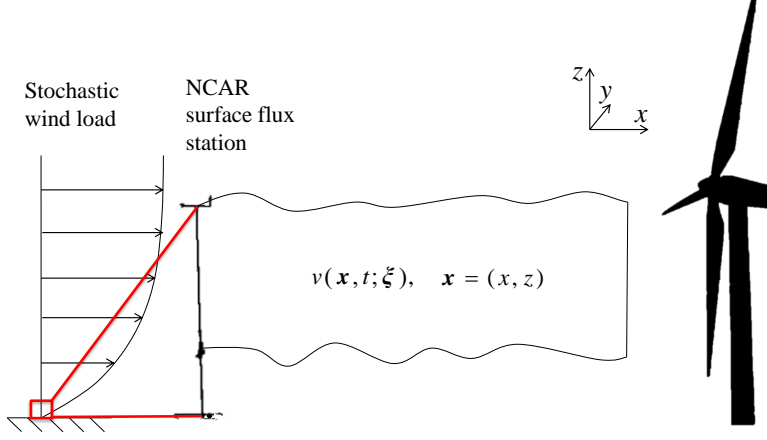


Figure 1: Setup of experiment

nally, one 2D snapshot taken at certain instant is constructed⁶. Similarly, snapshots of the wind field at other instants were created to represent one full day of measurements. We have data for 28 such days. The full meteorology data curated into this form can be represented as following matrix

$$\begin{matrix}
 \mathbf{v}_{11}(x, z) & \mathbf{v}_{12}(x, z) & \cdots & \mathbf{v}_{1m}(x, z) \\
 \mathbf{v}_{21}(x, z) & \mathbf{v}_{22}(x, z) & \cdots & \mathbf{v}_{2m}(x, z) \\
 \vdots & \vdots & & \vdots \\
 \mathbf{v}_{n1}(x, z) & \mathbf{v}_{n2}(x, z) & \cdots & \mathbf{v}_{nm}(x, z)
 \end{matrix}$$

In the matrix, each element $\{\mathbf{v}_{ij}(x, z), i = 1, 2, \dots, m; j = 1, 2, \dots, n\}$ represents a snapshot of wind field. Each column contains snapshots that span an entire day. Each row represents data for a particular interval measured over different days. Thus, each row of data can be considered to be *realizations* of the stochastic wind field at a particular interval.

It is noteworthy that the length of time interval that was used to obtain wind field snapshots can be arbitrarily chosen. Although, most guidelines for wind turbine design and siting suggest considering wind variabilities over a period of 10 minutes for wind data analysis [28], in this paper, analysis based

⁶Since no measurements were taken in the transverse direction, only 2D wind fields are analyzed in this work, though the mathematics is agnostic to dimensionality

on different choices of time intervals were performed and compared. In the results section, we will show that using 10 minutes time interval to construct wind field snapshots is an reasonable choice.

4. Algorithm & Implementation

The algorithmic details of the framework is outlined in Table 2. The meteorological data is provided in the common NetCDF (Network Common Data Form) format. The original meteorological data contains a variety of information, including wind speed and orientation, temperature, humidity, surface CO₂ flux. For this analysis we only consider the wind speed data. The wind speed data is first extracted from the NetCDF file using MATLAB. The matrix of wind flow snapshots is then constructed. The temporal covariance matrix is calculated from the snapshot matrix and stored as a datafile. Next, the eigenvalue problem (corresponding to the Bi-orthogonal Decomposition) is solved to get temporal and spatial-stochastic modes. Following this, the spatial-stochastic modes are decomposed into spatial functions and uncorrelated random variables via the Karhunen-Loève expansion. This is accomplished by solving an eigenvalue problem. The eigenvectors computed are the spatial functions and they are used to compute realizations of the uncorrelated random variables (via Eqn. 21). Finally, a Kernel Density Estimator is used to construct the PDFs of random variables ξ_j^i . At this time, all the necessary elements for constructing synthetic wind flow were known. Finally, synthetic wind realizations are constructed by reversing above process. The complete framework is implemented in C++. SLEPc [29], which is based on PETSc [30], was used in solving eigenvalue problems.

5. Results

As stated earlier, we use the meteorological data from the CWEX-11 experiment. The data consists of wind speed amplitude and direction measure at 1 Hz over 28 days. We will describe the stage wise construction of the low-dimensional model and its accuracy in the next few subsections.

5.1. BD results

As discussed previously, snapshots curated based on 10 minutes interval are used first. As a result, there are 144 snapshots in one day and we have 28 such days (samples). The mean component of the flow that is averaged

Table 2: Steps of computational framework.

-
1. Data preparation
 - (a) Extract wind speed data from NetCDF files
 - (b) Construct matrix of wind speed field snapshots
 2. Bi-orthogonal Decomposition
 - (a) Calculate temporal covariance from snapshots matrix
 - (b) Solve eigenvalue problem, get eigenvalues and eigenfunctions (temporal modes)
 - (c) Solve spatial stochastic modes
 3. Karhunen-Loève expansion
 - (a) Calculate covariance kernel of each spatial mode (M spatial modes in total)
 - (b) Solve eigenvalue problem for each spatial mode, get basis spatial functions
 - (c) Calculate observations of each random variable ξ_j^i based on Eqn. 21
 4. Kernel density estimation
 - (a) Estimate PDF for each random variable ξ_j^i based on observations
 5. Construct synthetic wind flow
 - (a) Get random samples from random variables ξ_j^i
 - (b) Construct stochastic spatial modes according to Eqn. 20
 - (c) Construct synthetic wind flow using Eqn. 4
-

across all 144 snapshots and 28 realizations according to Eqn. 2 is shown in Fig. 2. This plot shows us the mean wind profile, and clearly illustrates the vertical wind shear. Note that the x-axis represents length of the snapshots that are calculated based on 10 minute interval and average wind speed.

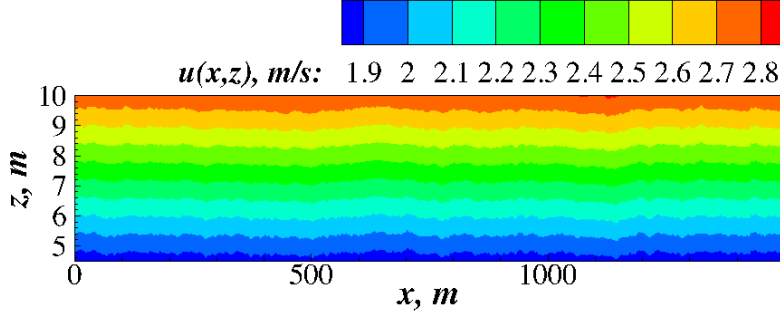


Figure 2: 10 minute period, averaged across all 144 time intervals and 28 realizations

The temporal covariance are calculated based on Eqn. 11. By taking the inner products across the spatial domain, we construct the covariance function $C(t, t')$ which is shown in Fig. 3. Note that C has block structure, with regions of high covariance (marked by the solid boxes) along the diagonal and regions of large negative covariance along the off-diagonal. This structure of the covariance function follows the dynamics of stable, and unstable stratification of the atmospheric boundary layer seen in the US central plains. We discuss this by dividing the analysis into several distinct time periods (marked by the solid boxes). Following meteorological practice, the data starts at Coordinated Universal Time (UTC, or Greenwich time) 00:00. The first period is UTC 00:00-06:00 (CST 18:00-00:00) which corresponds to the time between sunset to midnight. In this period, insolation and, thus, heating is gradually cut off and the temperature of atmosphere cools down (from the ground up) due to rapid cooling of the ground. Because of the heavier density of the cooler air, the cooler air stays at the bottom (close to the surface). This generates a stably stratified boundary layer that does not change during the duration of the night. This results in the high covariance between adjacent time periods marked in the lower left box. The second period is the region of reduced covariance between UTC 06:00-13:00 which is basically the time from midnight to shortly after sunrise. During summer, the nocturnal Great Plains low-level jet (LLJ) exists in Iowa. For a good overview of LLJ,

see Jiang et al. [31]. The LLJ causes low level turbulence and enhanced mixing which reduces the covariance between adjacent-time wind fields. The third time period is between UTC 13:00-00:00 which corresponds to sunrise and day time. Because of the sunrise, the ground is rapidly heated. As a result, the warmer air near the ground becomes buoyant and rises rapidly with its place being taken by compensating flow of colder air from higher up in the atmospheric boundary layer. Eddies are thereby created, changing from smaller scale to larger scales, finally becoming the circulatory motion that crosses the entire boundary layer so that the high speed free stream flow is brought in the circulation. This phenomenon usually happens at noon, which can also be found in Fig. 5 at UTC 18:00. In addition, the covariance function exhibits periodic behavior that is caused by the periodical motion of eddies.

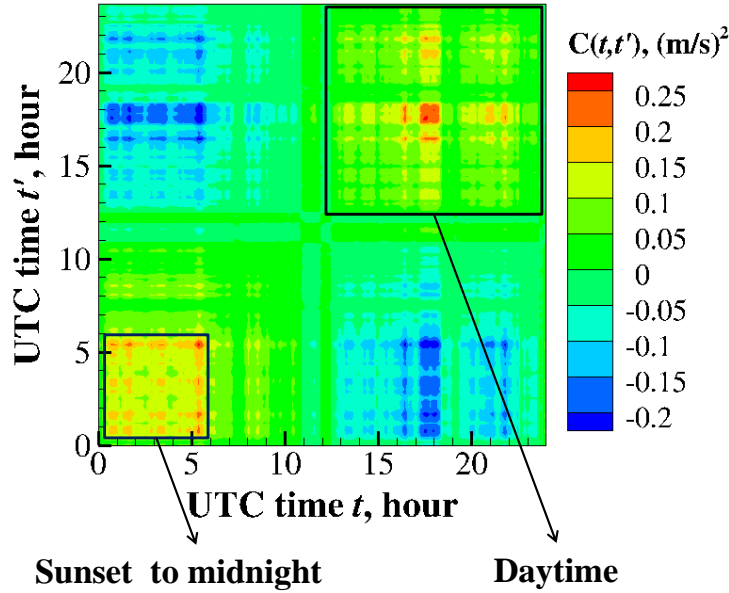


Figure 3: Covariance function $C(t, t')$

The temporal covariance function plays a very important role in this analysis since it provides almost all the needed information that describes the behavior of the wind flow. Compared to the original meteorological data, the temporal covariance is much easier to store and transmit, yet contains nearly the same amount of information as the original meteorological data. The

eigenvalues of the temporal covariance function are shown in Fig 4. Fig 4(a) compares the relative magnitudes of eigenvalues; the magnitudes of the eigenvalues provide a notion of how much energy about the data is stored in each spatial mode [32]. Notice that the first eigenvalue is much larger than the other subsequent eigenvalues, which means that the first mode contains the largest portion of the energy in the turbulence field. In Fig 4(b) this is represented as the cumulative fraction of energy contained in the first k modes. This plot provides us with a precise notion of how many terms are needed to incorporate [32], say, 90% of the information available in the data into the low-complexity model. The first five eigenvalues cover about 90 % of the total energy of the turbulence field, which ensures that a five term decomposition will have a 90% accuracy of representation. This illustrates the advantage of Bi-orthogonal Decomposition. As a reduced order model, Bi-orthogonal Decomposition is able to represent a random flow with much fewer random modes compared to the number of original snapshots. In other words, by using BD, we reduced the terms in representing the given stochastic random flow from 144 snapshots to a simple 5-variable parametrization. Fig. 5 shows

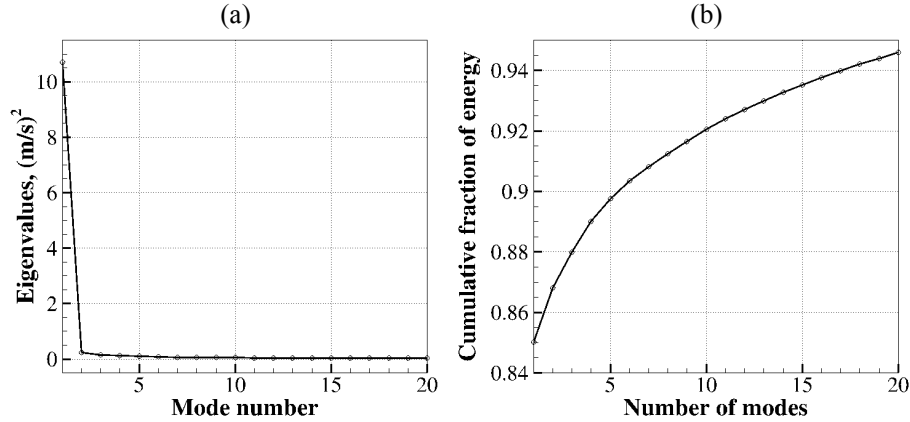


Figure 4: Spectrum of $C(t, t')$

the first three eigenfunctions of $C(t, t')$. Note that the eigenfunctions are the temporal functions T_i . They track how the fluctuation component of the random flow varies during a day. Based on the plot, the largest mean (and standard deviation, which is not reported here) of wind speed occurs at UTC 18:00 because of the reason stated previously. From Fig. 4, we notice that the first mode carries about 85% of the total energy in the turbulence. Therefore,

the trend of diurnal fluctuation of the turbulence field can be mostly seen in the first temporal mode, with the next two temporal modes look like white noise. Once eigenvalues and eigenvectors for temporal covariance function

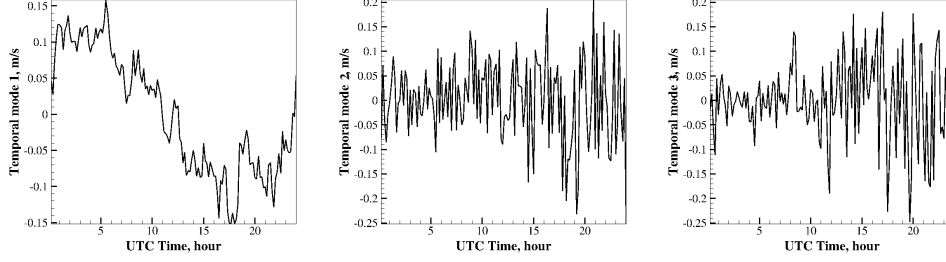


Figure 5: First three eigenvectors of temporal covariance function

are solved, the spatial-stochastic modes $a_i(\mathbf{x}, \boldsymbol{\xi})$ can be constructed (Eqn. 12 and Eqn. 14). Fig. 6 shows the expectation of the first three spatial modes. It is clear that the first spatial mode that carries the largest part of turbulence energy describes the vertical shear pattern, while the second mode describes the lateral shear pattern of the wind field. Higher modes that account for more complicated turbulence structures are insignificant since they contain limited turbulence energy.

5.2. KLE results

The spatial-stochastic modes are decomposed into deterministic modes and random variables using KLE. KLE results of the first three spatial modes are shown in Table 3. Covariance matrices, eigenvalues, deterministic spatial modes, and probably distributions of random variables are shown in the table from top to bottom. As discussed above, the distributions of ξ_j^i are estimated by using KDE.

From Fig. 6 we notice that as the mode order increases, the complexity of spatial mode also grows. As a result, higher order spatial modes require more KLE terms to be accurately represented. This statement can be verified by comparing results of the first three spatial modes. For instance, the smoothness of the covariance kernels, the complexity of deterministic modes, and the required numbers of KLE terms to achieve 90% of representation accuracies all increases as the mode number becomes higher.

It is worth to mention that the PDFs of ξ_j^i are not always Gaussian. Although Gaussian assumption is commonly used in turbulence analysis, non-

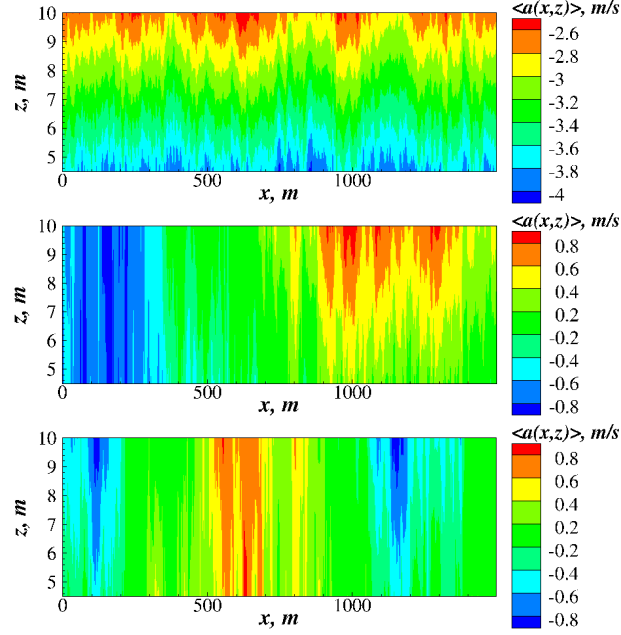


Figure 6: Stochastic spatial modes of $C(t, t')$

Gaussian variables do exist. Notice that each random sample corresponds to one day wind history, the randomness of this long term stochastic process may be different from the randomness of the more commonly used 10-minute average wind speed whose distribution is always regarded as Gaussian or Weibull.

5.3. The low-complexity wind model

The above two stages result in the calculation of the temporal functions, $T_i(t)$ (from Bi-orthogonal Decomposition), the spatial functions ($\mathbf{X}_j^i(\mathbf{x})$) (from KLE decomposition), and the uncorrelated random variables, ξ_j^i (from the KDE fitting).⁷ Putting it all together gives us the low-complexity model for the wind

$$\mathbf{u}(\mathbf{x}, t, \boldsymbol{\xi}) = \sum K_{ij} T_i(t) \mathbf{X}_j^i(\mathbf{x}) \xi_j^i \quad (25)$$

⁷The superscript in $\mathbf{X}_j^i(\mathbf{x})$ and ξ_j^i represents the index of Bi-orthogonal Decomposition terms whereas the subscript denotes the index of KLE terms.

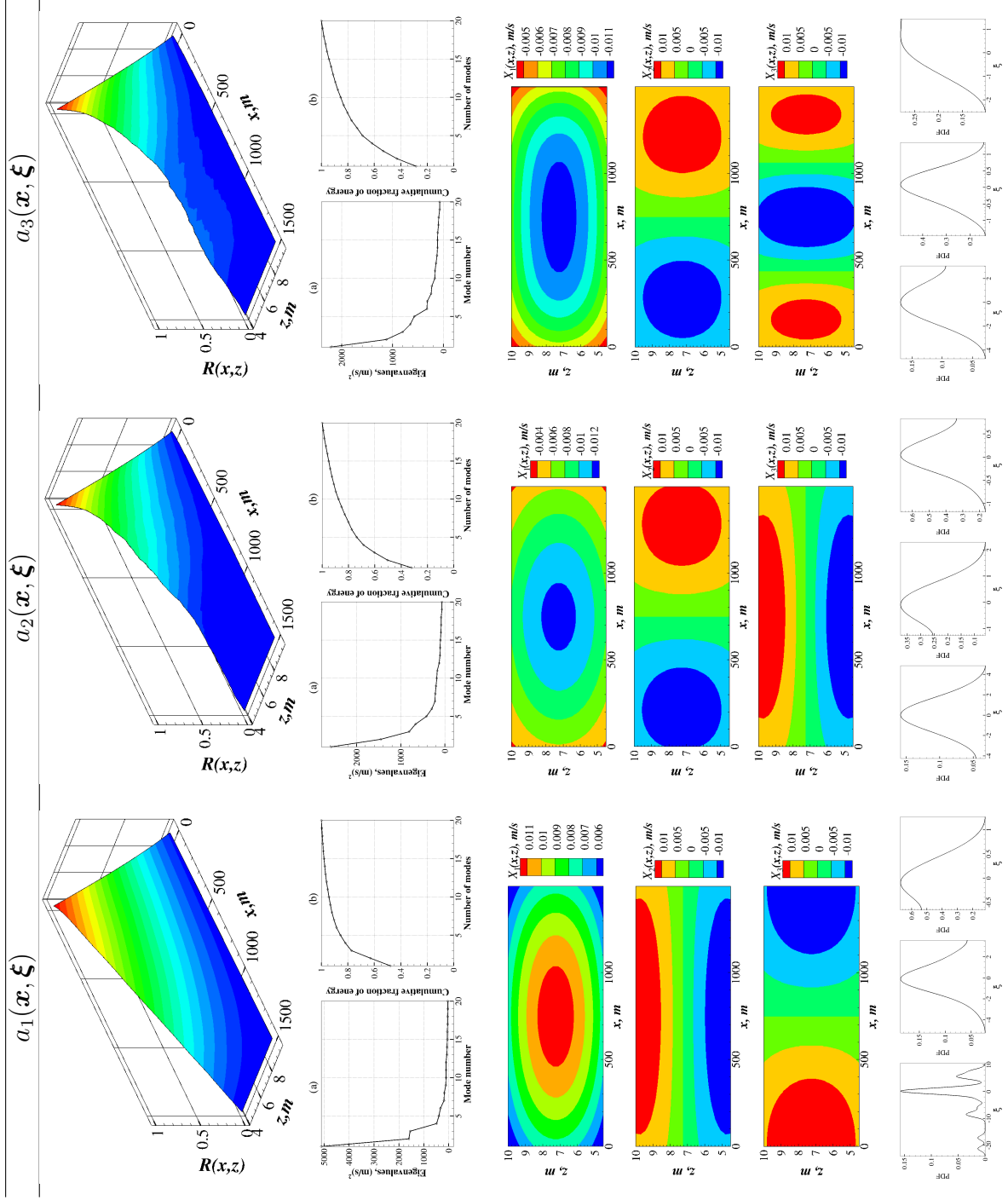


Table 3: KLE results of the first three spatial modes

Note that \mathbf{X}_j^i and T_i are deterministic functions that encode spatial and temporal correlations of the wind. Different realizations (or stochasticity) of the wind is included into the low-complexity model via the uncorrelated random variables, ξ_j^i . Note that the probability distributions of ξ_j^i are constructed in a data-driven way from the meteorology data. As we will show in the results section, only a few terms ($M = 3$) are required to reconstruct a synthetic wind snapshot that **contains all the temporal and spatial correlations exhibited by the original data**. Fig. 7 gives a graphical description of this process. Synthetic wind snapshots exactly mimicking the meteorological data can be constructed by sampling from the distributions of the random variables, ξ_j^i . Interestingly, only 800 KB of storage space is needed to store all the necessary information of the reduced-order model, whereas more than 200 MB of storage space is needed to store all the wind flow snapshots that are used in the analysis! This demonstrates the advantage of the low-complexity model in terms of data size.

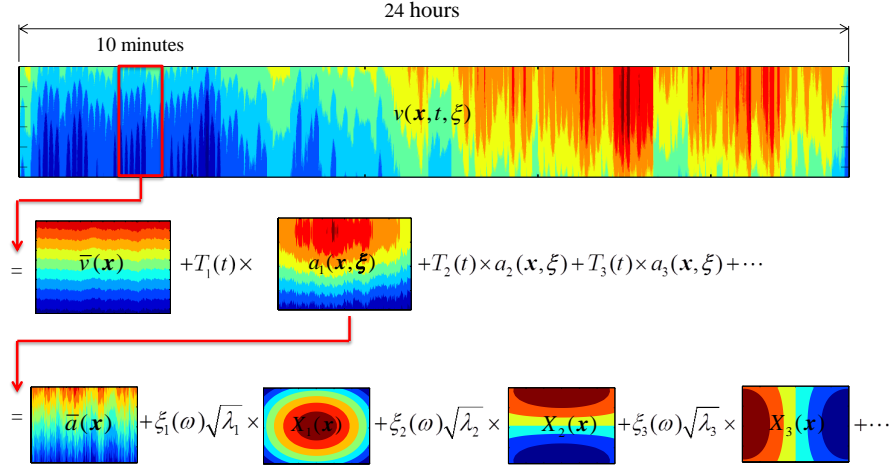


Figure 7: The low-complexity model: Process of constructing synthetic wind snapshots

5.4. Statistical comparison

As discussed previously, to get 90% representation accuracy, five terms in Bi-orthogonal Decomposition and seven terms in Karhunen-Loève Expansion are needed. However, for the purpose of demonstration, 1-, 3-, and 10-term

BD and three terms in KLE for each BD term were used in the analysis. In order to quantify the accuracy, 28 realizations (same as the number of samples of meteorological data set) of the synthetic wind flow are generated. Each realization is a 24-hour wind data consisting of 144 ten-minute snapshots (see Fig. 7 for one 24 hour synthetic dataset).

In Fig. 8, realizations of the stochastic flow at height 10m using one, three, and ten spatial modes are compared. Result shows that when smaller number of terms are used, the low frequency characteristics of the turbulence can be represented accurately. However, in order to describe high frequency behavior, higher modes in BD must be used. In the third plot that uses ten temporal modes, diurnal fluctuation in the variance of turbulence is addressed by including the second and third temporal modes (see Fig. 5).

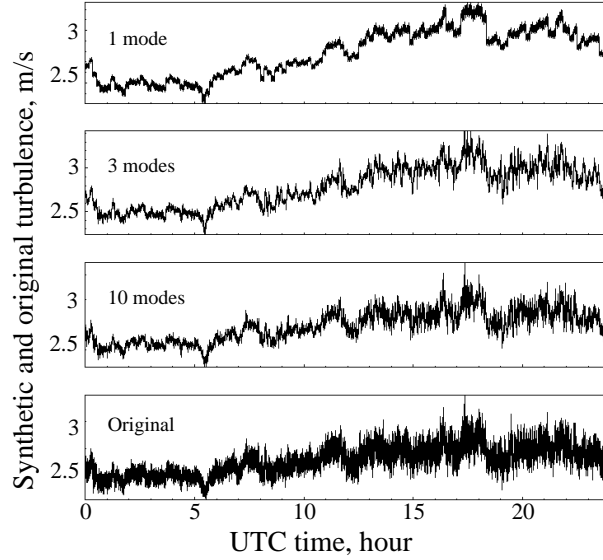


Figure 8: realizations of the stochastic flow

One of the most important variables is the power spectral density (PSD) that describes how the turbulent energy is distributed across the frequency spectrum. To verify the similarity of the reconstructed wind flow and the meteorological data, their PSD functions are compared in Fig. 9 (a). The figure shows that the synthetic wind flow accurately reproduces energy in low frequency region. However, more terms need to be used in the representation to preserve energy in high frequency region. It is worth mentioning that even

with only one BD mode, the synthetic wind mimics the true data in great detail.

Coherence spectrum is another important variable that describes the similarity of turbulence at two different spatial locations. Comparison of coherences between same set of time-series used in PSD analysis is given in Fig. 9 (b). Note that even by using only one BD mode, the synthetic wind can still preserve most of the spatial coherence information. This can be explained by the advantages of STD. Since STD performs orthogonal decompositions of both temporal and spatial covariances, the temporal variations and spatial correlations are preserved in an optimal way.

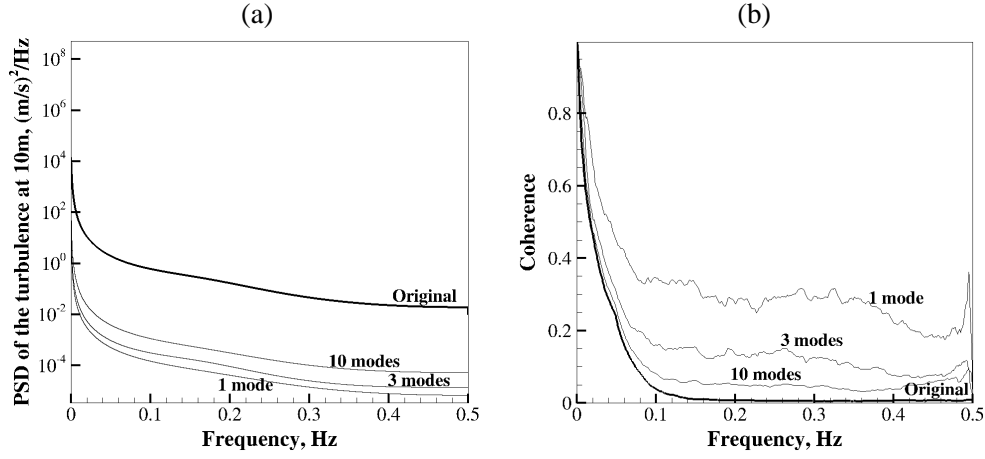


Figure 9: PSDs and coherences of the synthetic turbulence at 10 m using one, three, and ten modes compared with original flow

As previously mentioned, there are different choices of time intervals when constructing wind flow snapshots. In above content, only analysis on 10 minute snapshot was performed. The reason for choosing 10 minute snapshot can be demonstrated by comparing PSDs and coherences of the synthetic turbulence using different choices of time interval averaging processes. Fig. 10 shows results of such analysis. Since the accuracy of PSDs of synthetic wind only depends on how many total terms in the decomposition, there is no apparent difference in PSDs of different choices of time intervals. On the other hand, comparison of coherences tells us if performing analysis on snapshots that are constructed from less than 10 minute field, important information in turbulence coherence will be lost. Coherences correspond to

10, 15, and 20 minute snapshots are very close, which means the additional turbulence data provided by 15 and 20 minute snapshots, comparing to 10 minute snapshot, does not contain more coherence information. This result is consistent with GL guideline that states the coherence structure in turbulence can be assumed to be unchanged in 10 minutes period of time [28].

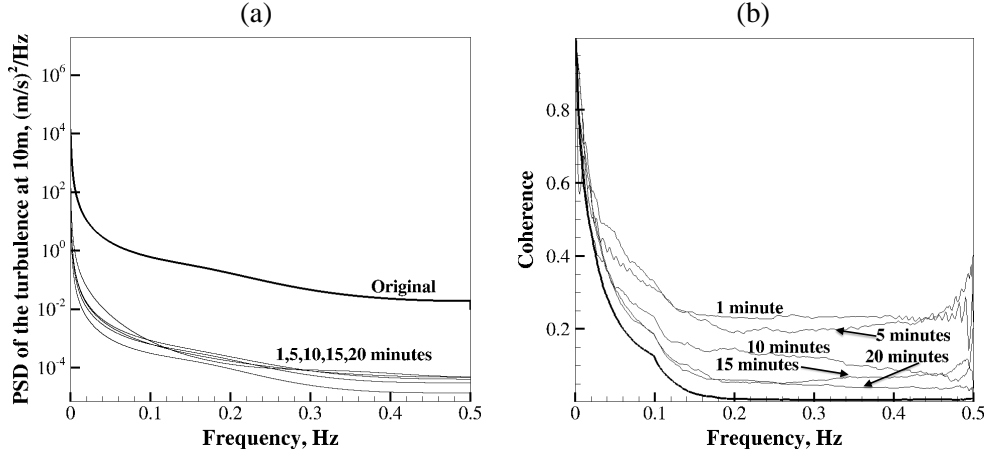


Figure 10: Comparing PSDs and coherences of the synthetic turbulence using different choices of snapshot

As illustrated in Fig. 4, a certain number of spatial modes are needed to achieve a desired representation accuracy. On the other hand, the number of random inputs to the reduced order model equals the number of spatial points on each snapshot. By using different choices of time intervals, the resulting snapshots have different dimensions. Table 4 shows the relationship between the number of random inputs and the number of needed spatial modes for certain level of accuracy. Based on the table, when the number of random inputs increases by 20 times, the needed random spatial modes to achieve 94% of accuracy only increases from 3 to 8. In other words, the required number of spatial modes does not strongly depend on the number of random inputs, which suggest that this approach (TSD) is practical for problems with very large stochastic dimensions.

The temporal covariance function of synthetic data $\hat{C}(t, t')$ is constructed. Comparing it with the temporal covariance of the original data reveals that they have almost identical pattern. The temporal error (or the information loss) can be defined as the L^2 -norm of the difference between the covariance

Table 4: Relationship between the number of random spatial modes (94% accuracy) and the number of random inputs.

Number of random inputs	1200	6000	12000	18000	24000
Number of spatial modes	3	5	6	7	8

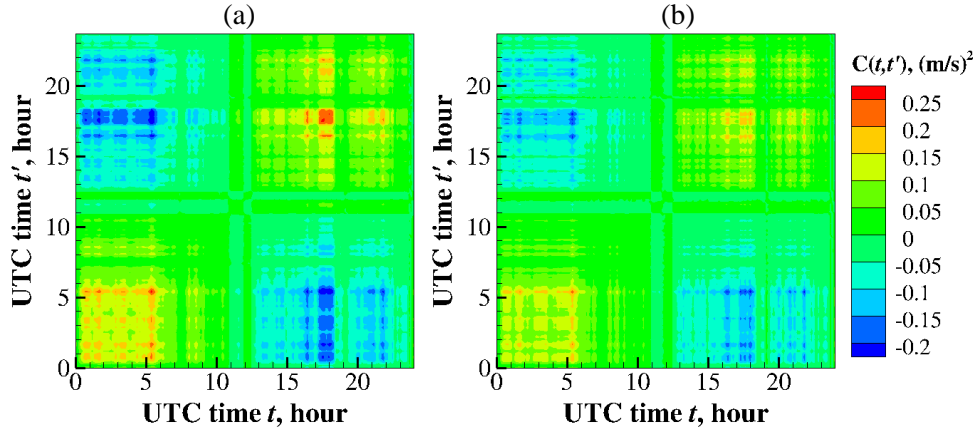


Figure 11: Comparison of covariance functions of original (a) and synthetic (b) flows

functions:

$$\epsilon = \frac{\|C - \hat{C}\|_2}{\|C\|_2} = \left(\frac{\sum_{i=1}^n \sum_{j=1}^n |C(t_i, t_j) - \hat{C}(t_i, t_j)|^2}{\sum_{i=1}^n \sum_{j=1}^n |C(t_i, t_j)|^2} \right)^{1/2}. \quad (26)$$

where $n = 144$ is the number of snapshots in 24 hours. The temporal information-loss using a 9-term expansion is 2.25%. Thus, a 9-term data-driven expansion reproduces the temporal and spatial covariance of the original meteorological data to 97.75% accuracy.

6. Discussion and Conclusion

Incorporating the effects of randomness in wind is critical for a variety of application involving wind energy. Continuous advances in data-sensing and meteorology has made possible the availability of large data-sets of location-, topography-, diurnal-, and seasonal sensitive meteorology data. While this data contains rich information, ease of use is bottlenecked by the unwieldy data-sizes. A pressing challenge is to utilize this data to construct a location-, topography-, diurnal-, and seasonal-, dependent low-complexity model that is easy to use and store. We formulate a data-driven mathematical framework that is capable of representing the spatial- and temporal- correlations as well as the inherent randomness of wind into a low-complexity parametrization. We leverage data-driven decomposition strategies like Bi-orthogonal Decomposition and Karhunen-Loève expansion for constructing the low-complexity model. We provide a software package that can be used to construct the low-complexity model to the community.

There are few points should be emphasized. First, the meteorological data used in this analysis is measured by only one met tower. Because the lack of the measurement on transverse direction, getting wind filed snapshots on the plane of turbine rotor becomes impossible. To circumvent this insufficiency in measurement, wind filed snapshots on the plane that is perpendicular to the rotor is used, for which certain time interval has to be chose to construct snapshots. It is worth noting that the framework is generally applicable to a variety of meteorology data, and the its applicability should not be affected by choosing different snapshot constructing time intervals. In addition, the framework is able to incorporate both short-term (10-minute) and long-term (years) temporal coherences as long as corresponding data is available. Third, while the mathematical framework developed here is used

to analyze wind speed, it can also be used to represent other atmospheric data such as temperature and carbon dioxide flux. This framework can also be naturally extended to represent ocean waves, which is crucial for off-shore wind turbine siting, layout and design analysis.

7. Acknowledgement

BG and QG gratefully acknowledge NSF 1149365.

8. References

- [1] J. C. Wyngaard, *Turbulence in the Atmosphere*, Cambridge University Press, 2010.
- [2] D. Xiu, G. E. Karniadakis, The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations, *SIAM Journal on Scientific Computing* 24 (2002) 619.
- [3] D. Xiu, G. E. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos, *Journal of Computational Physics* 187 (2003) 137–167.
- [4] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, *Journal of Computational Physics* 225 (2007) 652–685.
- [5] R. G. Ghanem, P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Dover Publications, 1998.
- [6] J. W. Larsen, R. Iwankiewicz, S. R. Nielsen, Nonlinear stochastic stability analysis of wind turbine wings by monte carlo simulations, *Probabilistic engineering mechanics* 22 (2007) 181–193.
- [7] D. C. Quarton, The Evolution of Wind Turbine Design Analysis-A Twenty Year Progress Review, *Wind Energy* 1 (1998) 5–24.
- [8] Y. Bazilevs, M.-C. Hsu, I. Akkerman, 3D simulation of wind turbine rotors at full scale. Part I: geometry modeling and aerodynamics, *Methods in Fluids* 65 (2011) 207–235.

- [9] Y. Bazilevs, M.-C. Hsu, J. Kiendl, 3D simulation of wind turbine rotors at full scale . Part II : Fluid structure interaction modeling with composite blades, *Methods in Fluids* 65 (2011) 236–253.
- [10] J. A. Carta, P. Pamírez, S. Velázquez, A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands, *Renewable and Sustainable Energy Reviews* 13 (2009) 933–955.
- [11] P. S. Veers, Three-dimensional wind simulation, Technical Report, Sandia National Labs., Albuquerque, NM (USA), 1988.
- [12] J. Mann, Wind field simulation, *Probabilistic engineering mechanics* 13 (1998) 269–282.
- [13] K. Saranyasoontorn, L. Manuel, On the propagation of uncertainty in inflow turbulence to wind turbine loads, *Journal of Wind Engineering and Industrial Aerodynamics* 96 (2008) 503–523.
- [14] B. J. Jonkman, TurbSim user’s guide: Version 1.50, National Renewable Energy Laboratory Colorado, 2009.
- [15] N. Kelley, B. Jonkman, G. Scott, J. Bialasiewicz, L. Redmond, The impact of coherent turbulence on wind turbine aeroelastic response and its simulation, in: *Windpower 2005 Conference Proceedings*.
- [16] J. Jimenez, Turbulent velocity fluctuations need not be gaussian, *Journal of Fluid Mechanics* 376 (1998) 139–147.
- [17] D. Venturi, X. Wan, G. E. Karniadakis, Stochastic low-dimensional modelling of a random laminar wake past a circular cylinder, *Journal of Fluid Mechanics* 606 (2008) 339–367.
- [18] D. Venturi, A fully symmetric nonlinear biorthogonal decomposition theory for random fields, *Physica D: Nonlinear Phenomena* 240 (2011) 415–425.
- [19] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener, *Mathematische Annalen* 63 (1907) 433–476.

- [20] J. Hunter, B. Nachtergaele, Applied Analysis, World Scientific Publishing Company, Singapore, 2005.
- [21] L. Mathelin, O. L. Maitre, Robust control of uncertain cylinder wake flows based on robust reduced order models, Computers & Fluids 38 (2009) 1168–1182.
- [22] D. W. Scott, Multivariate density estimation: theory, practice, and visualization, John Wiley & Sons Inc, 1992.
- [23] Q. Guo, Link of code used in current paper, 2012. [Http://www3.me.iastate.edu/bglab/pages/projects.html](http://www3.me.iastate.edu/bglab/pages/projects.html).
- [24] B. A. Turlach, Bandwidth Selection in Kernel Density Estimation: A Review, Technical Report, CORE and Institut de Statistique, 1993.
- [25] B. W. Silverman, Density estimation for statistics and data analysis, Chapman and Hall, London, 1986.
- [26] D. A. Rajewski, E. S. Takle, J. K. Lundquist, S. Oncle, J. H. Prueger, T. W. Horst, M. E. Rhodes, R. Pfeiffer, J. L. Hatfield, K. K. Spoth, R. K. Doorenbos, CWEX: Crop/Wind-energy EXperiment: Observations of surface-layer, boundary-layer and mesoscale interactions with a wind farm, 2012. [Http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-11-00240](http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-11-00240).
- [27] G. I. Taylor, The spectrum of turbulence, Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences 164 (1938) 476–490.
- [28] Germanischer Lloyd, Guideline for the Certification of Wind Turbines, Germanischer Lloyd, Hamburg, 2010 edition, 2010.
- [29] V. Hernandez, J. E. Roman, V. Vidal, SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems, ACM Transactions on Mathematical Software 31 (2005) 351–362.
- [30] S. Balay, J. Brown, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, H. Zhang, PETSc Web page, 2012. [Http://www.mcs.anl.gov/petsc](http://www.mcs.anl.gov/petsc).

- [31] X. Jiang, N.-C. Lau, I. M. Held, J. J. Ploshay, Mechanisms of the Great Plains Low-Level Jet as Simulated in an AGCM, *Journal of the Atmospheric Sciences* 64 (2007) 532–547.
- [32] P. Holmes, J. L. Lumley, G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, 1998.

9. Appendix

Appendix A Derivation of Minimizing Representation Error using Euler Lagrange Equation

Theorem 1. (Euler’s equation) *Let $J[y]$ be a functional of the form*

$$\int_a^b F(x, y, y') dx, \quad (\text{A.1})$$

defined on the set of functions $y(x)$ which have continuous first derivatives in $[a, b]$ and satisfy the boundary conditions $y(a) = A$, $y(b) = B$. Then a necessary condition for $J[y]$ to have an extremum for a given function $y(x)$ is that $y(x)$ satisfy Euler’s equation

$$F_y - \frac{d}{dx} F_{y'} = 0 \quad (\text{A.2})$$

□

The goal of BD is to minimize the error functional

$$\mathcal{E}[T_1, \dots, T_M] = \int_T \langle \varepsilon, \varepsilon \rangle_X dt. \quad (\text{A.3})$$

Substituting representation error (Eqn. 5) in above equation yields

$$\begin{aligned} & \mathcal{E}[T_1, \dots, T_M] \\ &= \int_T \left\langle u(\mathbf{x}, t, \boldsymbol{\xi}) - \sum_{i=1}^M K_i \boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi}) T_i(t), u(\mathbf{x}, t, \boldsymbol{\xi}) - \sum_{j=1}^M K_j \boldsymbol{\Phi}_j(\mathbf{x}, \boldsymbol{\xi}) T_j(t) \right\rangle_X dt \end{aligned} \quad (\text{A.4})$$

According to the definition of spatial inner product (Eqn. 7), the error functional becomes

$$\begin{aligned}
& \mathcal{E}[T_1, \dots, T_M] \\
&= \int_T \int_{\mathbf{X}} \left[u(\mathbf{x}, t) - \sum_{i=1}^M K_i \boldsymbol{\Phi}_i(\mathbf{x}) T_i(t) \right] \left[u(\mathbf{x}, t) - \sum_{j=1}^M K_j \boldsymbol{\Phi}_j(\mathbf{x}) T_j(t) \right] d\mathbf{x} dt \\
&= \int_T \int_{\mathbf{X}} u^2(\mathbf{x}, t) - 2u(\mathbf{x}, t) \sum_{i=1}^M K_i \boldsymbol{\Phi}_i(\mathbf{x}) T_i(t) \\
&\quad + \sum_{i=1}^M K_i^2 \boldsymbol{\Phi}_i^2(\mathbf{x}) T_i^2(t) d\mathbf{x} dt \\
&= \int_T \int_{\mathbf{X}} u^2(\mathbf{x}, t) - 2u(\mathbf{x}, t) \sum_{i=1}^M K_i \boldsymbol{\Phi}_i(\mathbf{x}) T_i(t) + \sum_{i=1}^M K_i^2 \boldsymbol{\Phi}_i^2(\mathbf{x}) T_i^2(t) d\mathbf{x} dt \\
&= \int_T \left[\int_{\mathbf{X}} u^2(\mathbf{x}, t) d\mathbf{x} - 2 \sum_{i=1}^M K_i T_i(t) \int_{\mathbf{X}} u(\mathbf{x}, t) \boldsymbol{\Phi}_i(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^M K_i^2 T_i^2(t) \right] dt \\
&= \int_T F(t, T_1, \dots, T_M, T'_1, \dots, T'_M) dt
\end{aligned} \tag{A.5}$$

where

$$\begin{aligned}
& F(t, T_1, \dots, T_M, T'_1, \dots, T'_M) \\
&= \int_{\mathbf{X}} u^2(\mathbf{x}, t) d\mathbf{x} - 2 \sum_{i=1}^M K_i T_i(t) \int_{\mathbf{X}} u(\mathbf{x}, t) \boldsymbol{\Phi}_i(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^M K_i^2 T_i^2(t)
\end{aligned} \tag{A.6}$$

In above derivation, the orthogonality of basis functions $\boldsymbol{\Phi}_i$ and T_i is applied. According to Theorem 1, the error functional $\mathcal{E}[T_1, \dots, T_M]$ has extremum when T_i satisfy Euler's equations

$$F_{T_i} - \frac{d}{dt} F_{T'_i} = 0. \tag{A.7}$$

That is

$$\int_{\mathbf{X}} u(\mathbf{x}, t) \boldsymbol{\Phi}_i(\mathbf{x}) d\mathbf{x} - K_i T_i(t) = 0, \tag{A.8}$$

which can be simplified as

$$T_i(t) = \frac{1}{K_i} \langle u(\mathbf{x}, t, \boldsymbol{\xi}), \boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi}) \rangle_X. \quad (\text{A.9})$$

Applying temporal inner product $\langle \cdot, T_i \rangle_T$ to Bi-orthogonal Decomposition (Eqn. 4) and considering the orthogonality of the temporal basis functions yields

$$\boldsymbol{\Phi}_i(\mathbf{x}, \boldsymbol{\xi}) = \frac{1}{K_i} \langle u(\mathbf{x}, t, \boldsymbol{\xi}), T_i(t) \rangle_T. \quad (\text{A.10})$$

Above two equations define a coupled relationship between T_i and $\boldsymbol{\Phi}_i$. Now that we have two unknowns and two equations, they can be solved by substituting Eqn. A.10 into Eqn. A.9, which results in eigenvalue problem for temporal modes

$$K_i^2 T_i(t) = \int_T C(t, t') T_i(t') dt', \quad (\text{A.11})$$

where $C(t, t')$ is called temporal covariance that can be obtained by taking the inner product in spatial domain, i.e

$$C(t, t') = \langle u(\mathbf{x}, t, \boldsymbol{\xi}), u(\mathbf{x}, t', \boldsymbol{\xi}) \rangle_X. \quad (\text{A.12})$$

Setting $\mu_i = K_i^2$, Eqn. A.11 can be transformed to

$$\mu_i T_i(t) = \int_T C(t, t') T_i(t') dt', \quad (\text{A.13})$$

where μ_i and $T_i(t)$ are eigenvalues and eigenfunctions of the covariance function $C(t, t')$. The optimal choice of temporal modes T_i and $\boldsymbol{\Phi}_i$ can be obtained by solving the eigenvalue problem.

Appendix B Numerical solution to the Generalized Eigenvalue Problem

We want to find the numerical solution of eigen-equation:

$$\int_{\mathbf{x}} C(\mathbf{x}_1, \mathbf{x}_2) f_i(\mathbf{x}_1) d\mathbf{x}_1 = \lambda_i f_i(\mathbf{x}_2). \quad (\text{B.1})$$

To this end, eigenfunction can be approximated by linear combination of N basis functions

$$f_k(\mathbf{x}) = \sum_{i=1}^N d_i^{(k)} h_i(\mathbf{x}). \quad (\text{B.2})$$

Substitute above equation to the eigen-equation and set the error to be orthogonal to each basis function yields

$$\sum_{i=1}^N d_i^{(k)} \left[\int_{\mathbf{X}} \left[\int_{\mathbf{X}} C(\mathbf{x}_1, \mathbf{x}_2) h_i(\mathbf{x}_2) d\mathbf{x}_2 \right] h_j(\mathbf{x}_1) d\mathbf{x}_1 - \lambda_n \int_{\mathbf{X}} h_i(\mathbf{x}) h_j(\mathbf{x}) d\mathbf{x} \right] = 0. \quad (\text{B.3})$$

Above equation can be written in matrices form

$$AD = BDA. \quad (\text{B.4})$$

$$A_{ij} = \int_{\mathbf{X}} \int_{\mathbf{X}} C(\mathbf{x}_1, \mathbf{x}_2) h_i(\mathbf{x}_1) h_j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2. \quad (\text{B.5})$$

$$B_{ij} = \int_{\mathbf{X}} h_i(\mathbf{x}) h_j(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{X}} H^T(\mathbf{x}) H(\mathbf{x}) d\mathbf{x}. \quad (\text{B.6})$$

$$D_{ij} = d_i^{(j)}. \quad (\text{B.7})$$

$$\Lambda_{ij} = \delta_{ij} \lambda_i. \quad (\text{B.8})$$

where $H(\mathbf{x}) = (h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_N(\mathbf{x}))$. Matrix A can be rewritten as

$$\begin{aligned} A &= \int_{\mathbf{X}} \int_{\mathbf{X}} H^T(\mathbf{x}_1) C(\mathbf{x}_1, \mathbf{x}_2) H(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \int_{\mathbf{X}} \int_{\mathbf{X}} H^T(\mathbf{x}_1) H(\mathbf{x}_1) C H^T(\mathbf{x}_2) H(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \int_{\mathbf{X}} H^T(\mathbf{x}_1) H(\mathbf{x}_1) d\mathbf{x}_1 C \int_{\mathbf{X}} H^T(\mathbf{x}_2) H(\mathbf{x}_2) d\mathbf{x}_2 \\ &= BCB. \end{aligned}$$

and

$$\begin{aligned} C(\mathbf{x}_k, \mathbf{x}_l) &= \sum_{i=1}^N \sum_{j=1}^N h_i(\mathbf{x}_k) C_{ij} h_j(\mathbf{x}_l) \\ &= h_k(\mathbf{x}_k) C_{kl} h_l(\mathbf{x}_l) \\ &= C_{kl}. \end{aligned}$$